White Paper Report

Report ID: 98480

Application Number: HD5046208

Project Director: H. Lewis Ulman (ulman.1@osu.edu)

Institution: Ohio State University Research Foundation

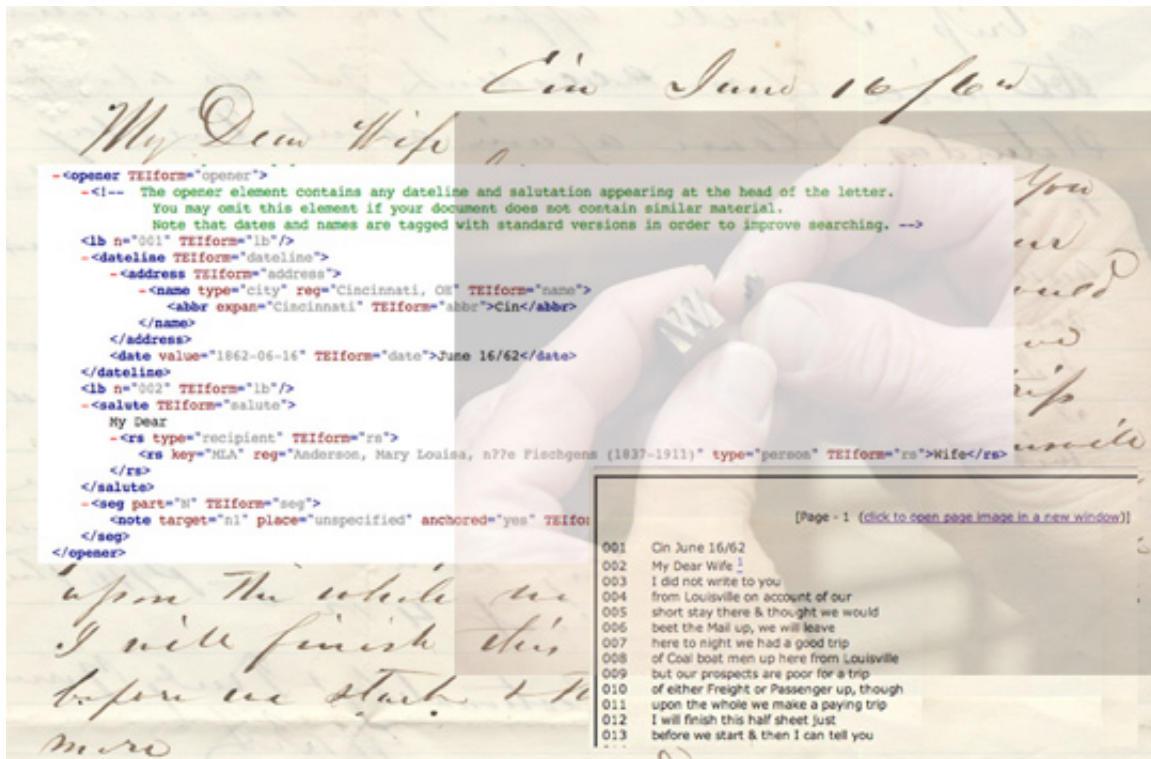Reporting Period: 1/1/2009-12/31/2009

Report Due: 3/30/2010

Date Submitted: 5/17/2010

White Paper
Digital Humanities Level II Start-up Grant #HD5046208

Reliable Witnesses

Integrating Multimedia, Distributed Electronic Textual Editions into
Library Collections and Preservation Efforts

Project Directors: H. Lewis Ulman (Department of English)
                            Melanie Schlosser (Ohio State University Libraries)

Project Team Members: Amy McCrory (Ohio State University Libraries)
                            Morag Boyd (Ohio State University Libraries)

The Ohio State University
17 May 2010

Table of Contents

Project Summary

This Level II Digital Humanities Start-up Grant developed a lifecycle model for preserving and sustaining complex electronic textual editions—editions that contain digital assets in multiple formats maintained on distributed systems supported by multiple organizations, and that provide users with multiple display options compiled on the fly in response to readers' actions (see Appendix 1). Scholars preparing printed textual editions for publication can reasonably expect that their work will be incorporated into a publisher's distribution network, made available through academic libraries, and preserved over the long term. An editor preparing an electronic textual edition can depend on no such tried and true conventions or networks for disseminating, cataloging, preserving, and sustaining access to digital humanities projects. We set out to provide scholars, library professionals, and IT professionals with workflows and best practices that would help them collaborate throughout the development of complex electronic editions and ensure that the digital assets comprising these projects—and their relationships to one another—can be preserved over the long term, and that the production environment of such projects can be sustained beyond the active involvement of the projects' creators.

Over the course of more than a year, we have interviewed faculty involved in digital humanities projects, library professionals, and information technology professionals; assessed the need for new practices adapted to digital preservation at our institution; and documented the resources and workflows currently available for, or adaptable to, long-term preservation of digital objects. We have also developed tools, institutional structures, and workflows for describing and archiving complex digital objects, as well as sustaining distributed digital production environments.

Preservation and Sustainability Tools and Workflows. Based on the data collected in our interviews, current literature in the field, and the characteristics of our case study projects, we have developed several assessment, preservation, and sustainability tools, including the following:

- a lifecycle map of complex digital projects that represents development and preservation milestones as interactions among scholars, library professionals, and IT professionals (see Appendix 2);

- a visual content manifest for complex digital projects that represents assets, the hardware on which those assets rely, and entities that enable the collaborative work of developing and preserving digital humanities projects (see Appendix 3);

- a semantic map that represents various presentations of the underlying digital assets from the end user's point of view (see Appendix 4);

- a Metadata Encoding and Transmission Standard (METS) profile for creating archival packages of complex digital projects (The Ohio State University Libraries Electronic Text Edition Archiving Profile: http://www.loc.gov/standards/mets/profiles/00000030.xml);

- a proposal for a Digital Humanities Network (DHN) designed to serve as a collaborative institutional curator for distributed digital projects, helping developers create preservation and sustainability plans and selecting, describing, and monitoring digital projects in production environments after the original developers are no longer involved with the projects;

Some of these tools and workflows will be easier to adapt to different projects, institutions, and cultural settings than others: for example, any library system should be able to adapt the METS profile to their needs, while our proposed Digital Humanities Network will serve mostly as a heuristic. We hope that these tools and workflows will initiate a fruitful conversation about how to build cultures of preservation for complex digital projects among scholars, librarians, and IT professionals in a variety of institutional settings.

Reliable Witnesses

Overview

Cultures of Preservation. Digital materials require cultures of description, preservation, and access every bit as robust as the practices and institutions that allow us to preserve manuscript and print materials. The devil of preservation—whether of print, digital, or other material artifacts—lies in the details of production, use, description, storage, conservation, and access. These concerns apply whether we are talking about acidic paper disintegrating on library shelves, digital files in obsolete formats, or media spread across computer systems whose links to one another have been broken. Preservation is further complicated by the distinction between preserving physical artifacts (books, manuscripts, floppy disks, flash drives) and preserving the information contained on those media in useful formats.

Contemporary, small-scale electronic textual editions are often multimodal, dynamic and distributed. In other words, they consist of multiple files and file types (e.g., XML, XSL, HTML, SQL, images) maintained on distributed hardware (which itself is supported by multiple organizations), and are often compiled for viewing on the fly in response to readers' actions. Preservation strategies for these project are necessarily complex. Anyone attempting to reconstruct such a project must to be able to locate all of the project's constituent parts and recognize what they are; the files must be in usable condition; and their contents need to be understandable to the people who want to use them. Moreover, if archived files are to be used in a manner similar to their original use, the files must be (1) compatible with current hardware and software, (2) translated into formats that are compatible with current hardware and software, or (3) used on reconstructed or emulated hardware and software that match the environment in which the project was originally developed. Small-scale digital humanities projects created by faculty often have the added vulnerability of relying solely on the developer's efforts and university computing accounts. Without a plan to transfer support when original developers are no longer involved in a project, many small-scale projects of potential value may be lost

The Specimen Case and the Garden. In a recent article entitled "Inkeeper at the Roach Motel," Dorothea Salo worries that focusing exclusively on preservation when designing institutional repositories leads to a situation in which documents are placed into repositories but never come out (Salo 2009). In this conceit, a "live" project gets placed in a repository and "dies" from lack of use. However, when attempting to preserve distributed, dynamic electronic textual editions, a somewhat different metaphor is needed. Like items in a specimen case, "live" digital projects must be "killed" before they are added to a conventional institutional repository such as DSpace, which cannot provide the sophisticated delivery systems needed by the project. In such applications, projects must be removed from the dynamic ecology of their production environments (the "garden") and frozen in a snapshot that is substantially different in appearance and functionality—a useful preservation strategy, to be sure, but not an approach to sustaining projects.

This project focuses on constructing preservation and sustainability models for small-scale, multi-format, distributed, dynamic electronic textual editions that would help developers transfer responsibility for projects, at their initiative, from the developers to institutional curators.

Major Activities

Working with case studies of distributed electronic textual editions under development at The Ohio State University (OSU), we worked over the course of the grant to devise local solutions to the problems posed by these editions and to articulate those solutions as general models or heuristics that could be adapted by other projects and institutions.

Needs Analysis: Interviews. To address the general issues outlined above, the project team conducted seventeen interviews over a two-month period with faculty scholars, library professionals, and IT professionals at OSU. The purpose of the interviews was to: (1) document the resources available at OSU for preserving and sustaining complex electronic textual editions; (2) understand how members of different units perceived their units' roles in the preservation and sustainability of complex digital projects; and (3) chart the necessary steps our university community would have to take in order to preserve digital assets and sustain digital projects effectively.

During our conversations with library and IT professionals, we asked them how they understood the roles or missions of their units. When speaking with librarians, we also asked what materials or types of materials constituted the Libraries' "collection." This question was an attempt to understand how items enter the collection, and how our case studies might fit into it. With all interviewees, we also reviewed drafts of our lifecycle model as well as a visual content manifest and a semantic map of a case study project being developed by one of the project team members (see descriptions below).

After each interview, the interviewers summarized their notes and submitted the summaries to interviewees for approval or, if necessary, correction and revision. Once approved, the summaries were shared with external consultants and formed the basis of the project team's needs analysis and environment scan.

Lifecycle Model. To help us understand the lifecycle of electronic textual editions, we developed a draft model and asked interviewees to "place" their units' roles and their individual roles on the model, and to comment on the tentative roles that we assigned to their cohort (faculty scholars, library professionals, and IT professionals). Interviewees helped us refine the model and better understand the need for consultation among various units throughout the lifecycle (see Appendix 2).

Visual Models of a Case Study Edition. To help interviewees people understand the issues involved in preserving and sustaining "multimedia, distributed, dynamic electronic textual editions," we developed two visual models. The models help explain the nature of such projects and their relationship to various library and IT units and project staff. The visual models also facilitated communication between the project director editing one of our case studies (Ulman) and the project director who was developing a metadata/preservation model for the project (Schlosser). (The case study edition, "Samuel Sullivan Cox's 'Journal of a Tour to Europe' (1851)," may be viewed online at http://people.cohums.ohio-state.edu/ulman1/SSCoxJournal/).

One model—the *visual content manifest*—represents all of the file types that constitute the edition; the various servers on which those files are stored, processed, or created on the fly; and the organizations that manage those servers (see Appendix 3). The other model—a *semantic map*—represents the semantic content of the case study edition from the end user's point of view. Similar to a table of contents, this view described the major semantic components of the edition, including different views or versions of the underlying XML transcription as well as contextual materials. It also references the visual content manifest to show how different systems and types of files contribute to each part of the edition. This model also informed the construction of a corresponding section of the preservation document for the project (see Appendix 4).

Needs Assessment and Environment Scan. When all interviews were completed, the project directors analyzed the summaries and composed a needs assessment and environment scan, which we sent—along with the original project proposal, summaries of individual interviews, and the interview materials—to our consultants.

Consultant Visits. In December 2008 and March 2009 our consultants visited OSU to review the project and meet with OSU faculty, library professionals, and IT professionals.

In December, Laura Mandell (Miami University of Ohio) and Peter Robinson (University of Birmingham, England) met with the project team, the OSU Libraries Executive Committee, faculty involved in digital media projects, and IT staff from several units on campus that support our case study projects. They also visited several support facilities, including the OSU Libraries Conservation Department and Digitization Lab, and the College of Humanities media production studio. In March, Oya Rieger (Cornell University) met with the project team, OSU Libraries' Assistant Director for Information Technology and Technical Services, the Head and staff of the Libraries' Scholarly Resources Integration Department, and staff associated with the Knowledge Bank, OSU's institutional repository. After their visits, each consultant submitted a written report to the project directors.

Metadata Encoding and Transmission Standard (METS) Profile and Document. Once the initial round of interviews and consultant visits was complete, a significant amount of time was devoted to creating a Metadata Encoding and Transmission Standard (METS) Profile to guide the description of the case study project. Profiles are a standardized way of providing guidance to those engaged in creating similar METS documents, and are collected and registered by the Library of Congress.

A preservation plan is a crucial element of our lifecycle model for distributed projects, and we felt that we could best take advantage of the preservation infrastructure available in the Libraries by creating an archival version of the project. The METS standard was chosen for this purpose because of its flexibility and broad

adoption in libraries for describing complex digital objects. DSpace, the software behind OSU's institutional repository, does not natively handle complex, dynamic, multimedia projects. In addition, its native Dublin Core (DC) metadata cannot capture adequate descriptive, administrative, and structural information to preserve such projects. Once the METS profile and test METS document were complete, we worked with the Knowledge Bank team in the Libraries to determine the best way to submit and present an archival package of the test project. In addition, the METS Profile was submitted to the Library of Congress, and it has since been formally registered as The Ohio State University Libraries Electronic Text Edition Archiving Profile (http://www.loc.gov/standards/mets/profiles/00000030.xml)

Digital Humanities Network. If the METS document described above offers a "specimen case" in which to *preserve* the assets comprising complex digital projects (and their relationships to one another) at any stage of their development, *sustaining* such projects in active production environments requires collaboration among faculty, IT professionals, and library professionals. To meet that need, we have developed a model for a Digital Humanities Network (DHN), which we are currently discussing with OSU Libraries and various other stakeholders, including IT service providers and academic administrators. The DHN would be "housed" in the Libraries because of its unique role in acquiring, describing, providing access to, and preserving cultural heritage materials over the long term.

As we have envisioned it, the DHN would consist of a standing committee within OSU Libraries, co-chaired by a library professional and an academic faculty member appointed by the Director of Libraries. Other members would come from OSU libraries, distributed IT support staff, and Humanities faculty. The DHN's function would be to *connect* faculty, students, librarians, and IT professionals who are creating or supporting digital humanities projects; *support* their efforts to develop robust, sustainable projects; and *curate* a set of completed projects that meet specific criteria, sustaining them beyond the involvement of their original creators.

Continuing Work

Preservation. Once the initial case study project has been successfully archived in the Knowledge Bank (OSU's institutional repository), we will begin archiving additional projects and share the workflow with other projects on campus.

Sustainability. If established as currently proposed, the DHN will be a standing committee of the OSU Libraries drawing its members from OSU Libraries staff, distributed IT professionals, and faculty researchers. It will work closely with our Office of the Chief Information Officer and Executive Dean for the College of Arts and Sciences to coordinate its efforts to preserve and sustain innovative digital humanities scholarship with the academic mission of the University.

Recommendations

Our preservation and sustainability planning materials are intended only as heuristic models; individual projects will need to adapt them, and the strategies that inform them, to local circumstances. Further, the materials are most relevant to small-scale projects without ongoing funding that must share computing infrastructure—e.g., Web servers, XML publishing systems, image and media servers, databases—with other systems and applications. That said, our informants (library professionals, IT professionals, and faculty researchers) stressed the following preservation and sustainability recommendations:

- Involve library professionals *throughout the lifecycle* of digital humanities projects. They can help project directors consider the preservation implications of decisions about digitization, file formats, and delivery systems in light of local resources and public standards.

- In collaboration with IT and library professionals, discuss plans to preserve a project's digital assets *and* sustain the project's production environment. Consider preserving snapshots of a project's digital assets—and documenting their relationships to one another—at significant stages in development (e.g., initial public release, major revisions). Consult with library professionals early on about options for transferring curatorial responsibility when the original developers are no longer involved in the project.

- Develop preservation and sustainability plans tied to *institutional roles and systems* rather than individual accounts or relationships.

- Describe projects fully. In collaboration with library professionals, IT support staff, and project staff, develop representations of a project's *lifecycle*, *digital assets*, and *semantic contents* that speak clearly to each group's needs. Digital humanities projects look very different to system administrators, metadata librarians, project staff, and end users. Robust plans for preservation and sustainability require that all of those perspectives integrate effectively.

Resources

METS Profile: Metadata Encoding and Transmission Standard (METS) profile for creating archival packages of complex digital projects (The Ohio State University Libraries Electronic Text Edition Archiving Profile: http://www.loc.gov/standards/mets/profiles/00000030.xml).

Web Site: Current versions of our preservation and sustainability planning templates can be viewed and downloaded from https://digitalhumanities.osu.edu/.

APPENDICES

Appendix 1: Conceptual Model of a Distributed Electronic Textual Edition
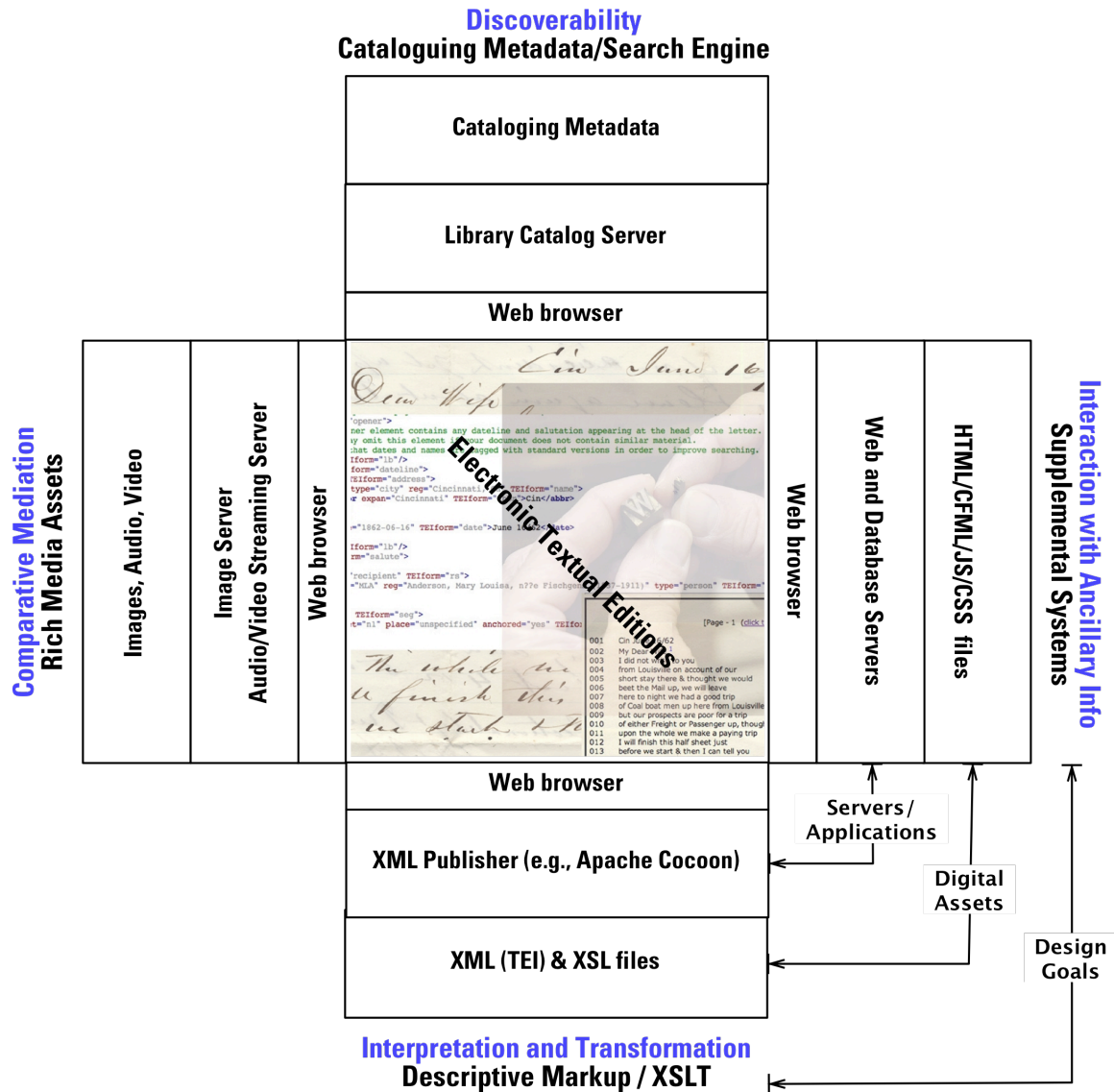
Conceptual model of a complex electronic textual edition, illustrating how multiple digital assets, systems, and processes converge in the Web browser to constitute a multimodal, distributed, dynamic electronic textual edition.

Appendix 2: Life-cycle Model for Preservation and Sustainability

Version 2.0 • 2008-11-24

A Life-Cycle Model For Integrating Distributed Electronic Textual Editions Into Library Collections and Digital Preservation Efforts

Main lifecycle stages (center flow): Authorize → Design → Develop → Credential → Acquire → Describe/Catalog → Provide Access → Preserve

- Authorize: Publishing Permissions — Lib / Edit negotiate
- Design: IT SLA(S) for Development — IT / Edit negotiate; Standards Best practices — Lib / Std / Edit (establish, consults, consults)
- Develop: Archival Images (TIFF) — Lib produces; All Other Digital Assets — Edit produces
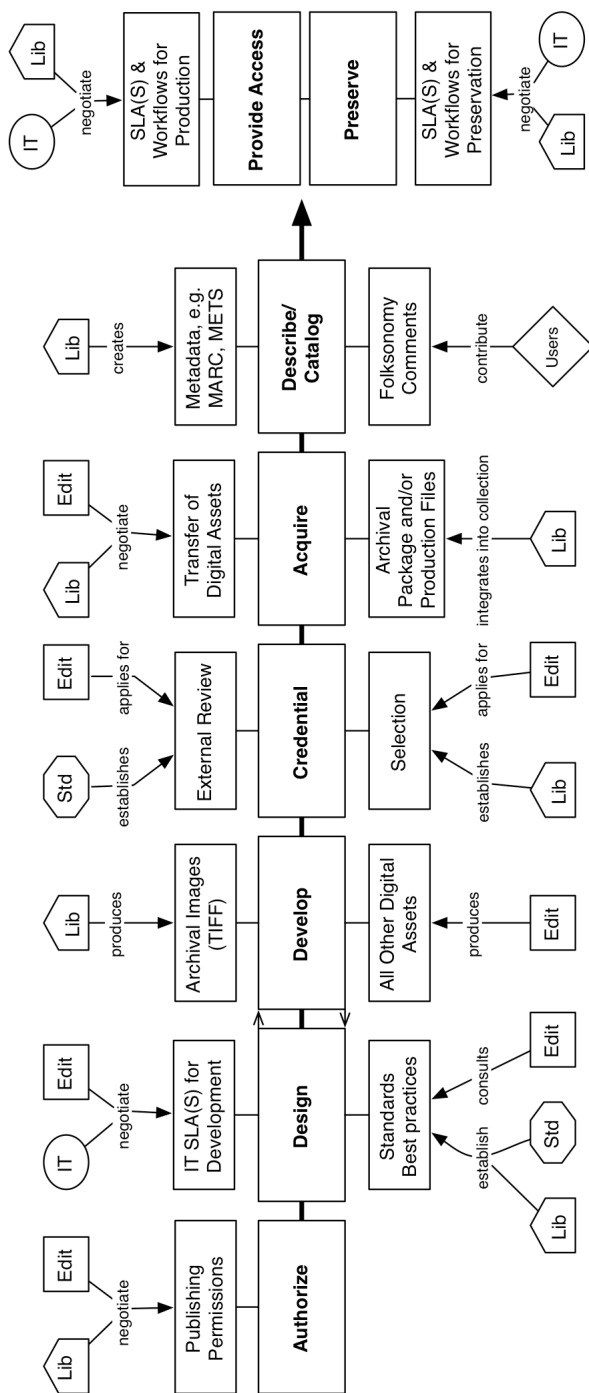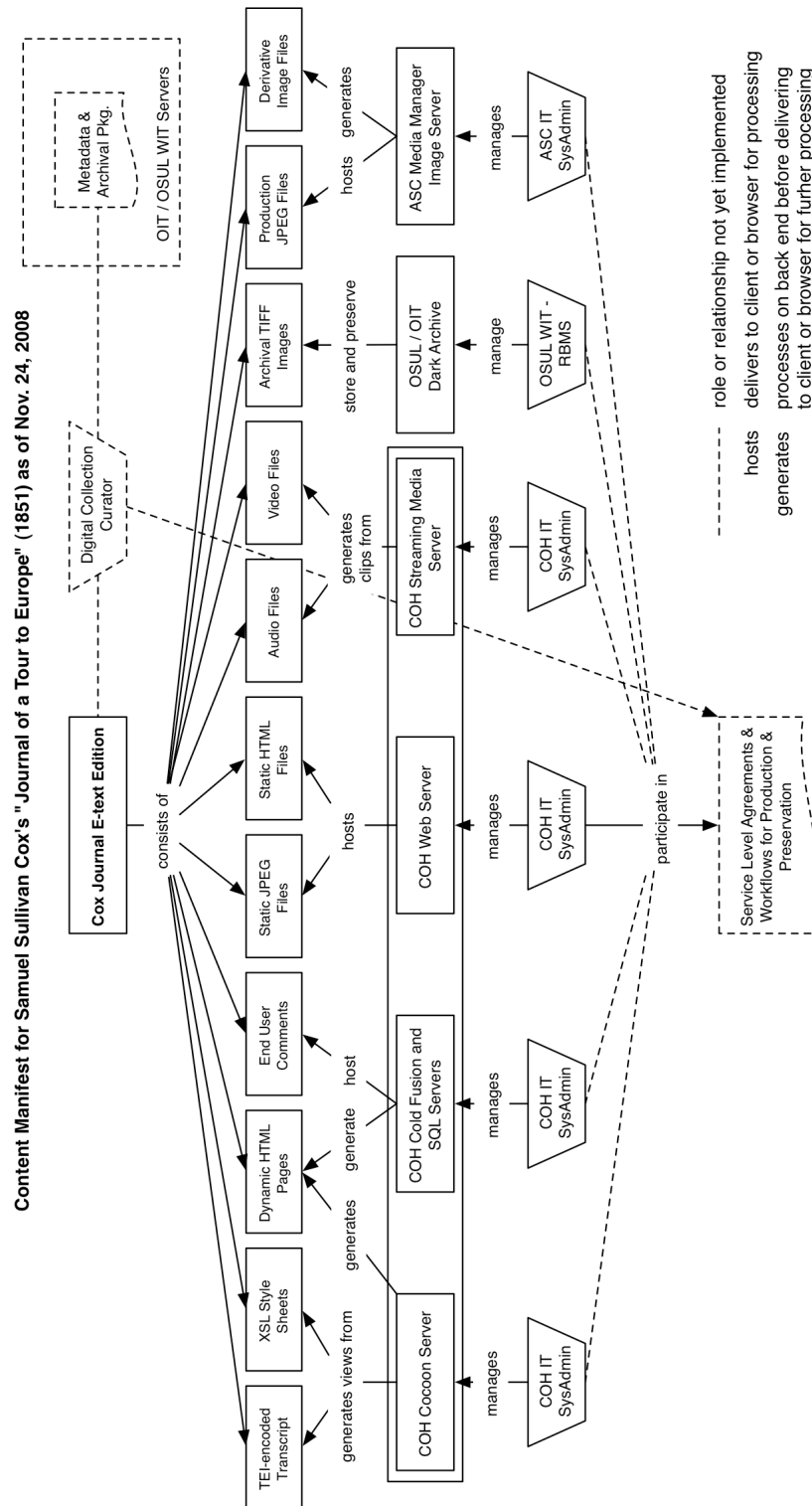- Credential: External Review — Std establishes, Edit applies for; Selection — Lib establishes, Edit applies for
- Acquire: Transfer of Digital Assets — Lib / Edit negotiate; Archival Package and/or Production Files — Lib integrates into collection
- Describe/Catalog: Metadata, e.g. MARC, METS — Lib creates; Folksonomy Comments — Users contribute
- Provide Access: SLA(S) & Workflows for Production — IT / Lib negotiate
- Preserve: SLA(S) & Workflows for Preservation — IT / Lib negotiate

Key Questions
What is archival? What is ephemeral?
What is the *ideal* workflow at this stage (whether feasible or not)?
What is the analogous treatment of traditional artifacts?
What plans are in place for technological obsolescence?
What happens if any individual, at any stage, is "abducted by aliens"?
How does your unit/role fit into this workflow?
Who else do you think we should interview?

Key Parties
Edit - Editors and Project Staff
Lib - Library Professionals and Admin.
IT - IT Professionals and Admin.
User - End Users
Std - Standards and Academic Orgs.

Key Features of E-Text Projects
Multiformat / Multimedia
Interactive / Dynamic
Collaborative
Distributed Across IT Systems
Recomposable (e.g., XML)
Ongoing
Open Access

Workflows
Roles
Actions
Objects
Triggers
Sequences

Content Manifest for Samuel Sullivan Cox's "Journal of a Tour to Europe" (1851) as of Nov. 24, 2008

Semantic Content Map for Samuel Sullivan Cox's "Journal of a Tour to Europe" (1851) as of Nov. 24, 2008

```
                              Cox Journal E-text Edition

        Static HTML, CFM,            consists of              XSL Transformations
     Javascript; MM Images:                               of XML Source Doc; MM Images;
         Streaming Media                                        Streaming Media
                              XSL Transformations
                           of XML Source Doc; MM images
```

Project Web Site

- Home Page
- Index to MSS Images
- Restoration Documentary
- Maps
- Markup Guidelines
- Guide to Cox's Hand
- Acknowledgements
- Comments and Corrections
- About the Editors

Editorial Introduction

- About the Source Document
- About the Electronic Edition
- Revision History

Views of the Journal

Reading View	Diplomatic View	MSS Image / Text View	Audio Version
Front Cover	Front Cover	Front Cover	Journal Entries
Front Endpapers	Front Endpapers	Front Endpapers	Portrait of S. S. Cox
Title Page	Title Page	Title Page	
Journal Entries (Dateline, Text, Illus.)	MSS pages / Lines	MSS pages / Lines / Page Images	Links to Zoomable MSS Page Images in Flash Player
Back Endpapers	Back Endpapers	Back Endpapers	
Back Cover	Back Cover	Back Cover	
Explanatory Annotations	Explanatory Annotations	Explanatory Annotations	
Works Cited List	Textual Annotations	Textual Annotations	
	Works Cited List	Works Cited List	